

Guide d'utilisation de *SyllabO+*

Version du 20 janvier 2018



LABORATOIRE DES NEUROSCIENCES
DE LA PAROLE ET DE L'AUDITION

SPEECH AND HEARING
NEUROSCIENCE LABORATORY

Guide d'utilisation

Avant de télécharger la base de données entière, il est recommandé d'avoir une bonne idée de l'information dont vous avez besoin, car les possibilités sont nombreuses et certaines de nos bases de données sont assez volumineuses (25 MB pour la base de données des syllabes)! *SyllabO+* contient en effet une grande quantité de données et il est possible qu'une partie seulement soit pertinente pour vos objectifs.

Instructions

1. Allez à <http://syllabo.speechneurolab.ca/> et après avoir répondu à quelques brèves questions, vous aurez accès aux **bases de données *SyllabO+***. Vos réponses demeureront strictement confidentielles et ne seront utilisées qu'à des fins de rapports pour les agences de financements et autres partenaires du projet afin de quantifier l'utilisation de nos bases de données.

Accéder à / Go to SyllabO+

2. Vous pourrez alors **choisir** entre **télécharger une base de données complète** (informations extraites à partir de tous les enregistrements audio) – que ce soit pour les **phones**, pour les **syllabes**, pour les **mots** ou pour les **lemmes** – ou **parcourir en ligne et télécharger une base de données spécifique**. Les bases de données complètes contiennent les statistiques d'utilisation globales, c'est-à-dire pour tous les contextes et tous les locuteurs (peu importe leur genre et leur âge). Elles représentent ainsi l'usage général des locuteurs adultes du français québécois en terme de syllabes, séquences de syllabes et phones.

L'image ci-dessous illustre la base de données complète des syllabes.

Base de données des syllabes

Lorsque vous téléchargez toute la base de données de syllabes, trois fichiers sont inclus dans un document compressé (.ZIP): soit les syllabes, les paires et les triades.

Des versions alternatives de la base de données sont également disponibles ci-dessous aux formats .XLSX (avec les caractères spéciaux API déjà encodés) et au format .CSV sans caractères spéciaux (transcription avec l'alphabet [SAMPA](#) – Speech Assessment Methods Phonetic Alphabet).

[Télécharger la base de données de SYLLABES](#)

Vous pouvez également choisir de parcourir et de télécharger une **base de données spécifique** en cliquant sur le bouton suivant :

[Pour faire une recherche en ligne, cliquez ici](#)

La base de données contiendra alors les statistiques d'utilisation extraites à partir d'un sous-ensemble des enregistrements audio. Les statistiques d'utilisation seront alors compilées en fonction des paramètres suivants, individuellement ou combinés: contexte de communication (formel, informel), âge, et genre des locuteurs. Par exemple, il est ainsi possible de télécharger les statistiques d'utilisation des mots seulement pour les jeunes adultes, ou pour les femmes en contexte formel, etc. Il est également possible d'afficher en ligne les résultats de la recherche.

Voici le formulaire pour parcourir en ligne ou pour télécharger la base de données désirée :

Type*

Contexte d'élocution - *Communication context*

Genre des locuteurs - *Speaker gender*

Résultats par page - *Results per page**

Âge minimum - *Minimum age*

Âge maximum - *Maximum age*

[Voir en ligne - Browse online](#)

[Exporter en CSV* - Export as CSV*](#)

3. Si vous avez **téléchargé une base de données** complète, vous obtiendrez plusieurs fichiers différents, compressés en format ZIP. Pour les syllabes, vous obtiendrez un fichier pour les syllabes individuelles et leurs statistiques, un pour les paires de syllabes et leurs statistiques et un pour les groupes de 3 syllabes (triades) et leurs statistiques. Pour les phones, vous obtiendrez un fichier pour les phones individuels et leurs statistiques et un pour les paires de phones (diphones) et leurs statistiques. Pour les mots ou lemmes, vous obtiendrez un fichier pour les mots/lemmes individuels et leurs statistiques, un pour les paires de mots/lemmes et leurs statistiques et un pour les groupes de 3 mots/lemmes (triades) et leurs statistiques.
4. Si vous avez **téléchargé une base de données spécifique**, vous obtiendrez seulement le fichier sélectionné (unités seules, paires d'unités ou triades d'unités).
5. Tous les fichiers sont au **format CSV** (Comma Separated Values), qui est une manière simple de stocker des données tabulaires en fichiers textes bruts. Dans les fichiers, les unités sont transcrites en Alphabet Phonétique International (API) (pour phones et syllabes seulement, non applicable pour mots et lemmes). À noter que deux autres formats sont maintenant disponibles : format XLSX avec les caractères spéciaux API déjà encodés ainsi que format .CSV sans caractères spéciaux (transcription avec l'alphabet SAMPA – Speech Assessment Methods Phonetic Alphabet). Ces formats ne sont toutefois disponibles que pour le téléchargement des bases de données complètes.
6. Si vous téléchargez les fichiers .CSV (API), le plus simple est de les ouvrir avec le logiciel **LibreOffice** (tableur). Nous recommandons l'utilisation de ce logiciel parce qu'il permet de traiter les caractères phonétiques (API) correctement. Microsoft Excel ne traite pas ces caractères spéciaux correctement (à moins d'utiliser les fichiers déjà en format XLSX). À noter:

LibreOffice est un logiciel gratuit et Open source, disponible pour téléchargement à <http://fr.libreoffice.org/> . Pour la version en français, téléchargez l'installateur principal, puis l'interface en français. (Allez à <http://www.libreoffice.org/> pour la version en anglais).

Installateur principal

Sélectionné : LibreOffice 5.0.5 pour Mac OS X x86_64 (10.8 ou plus récent requis) - modifier ?

TÉLÉCHARGER LA VERSION 5.0.5

193 MB (Torrent, Informations)

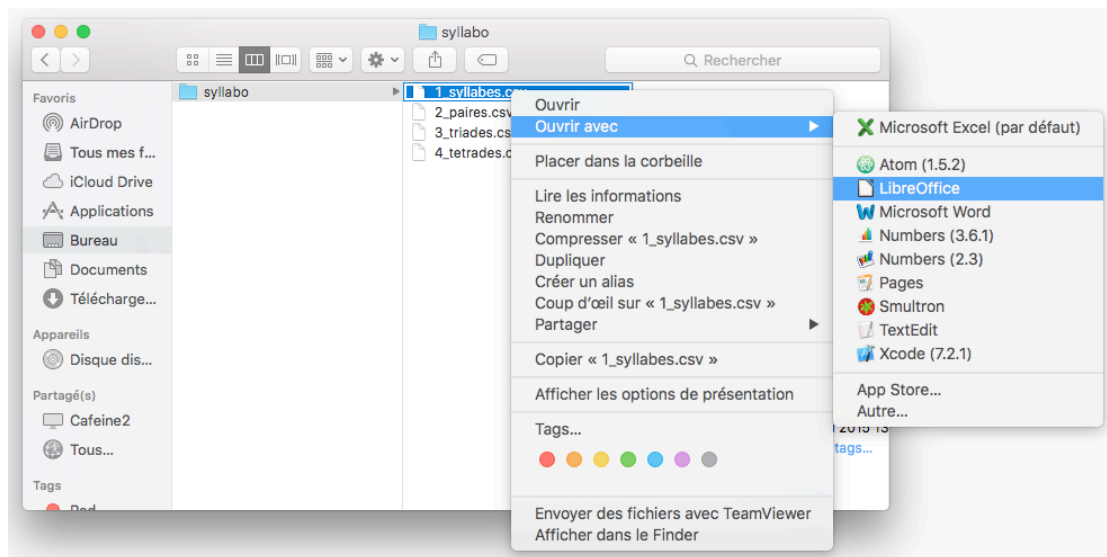
Obtenir l'interface LibreOffice traduite en français

vous souhaitez une autre langue ?

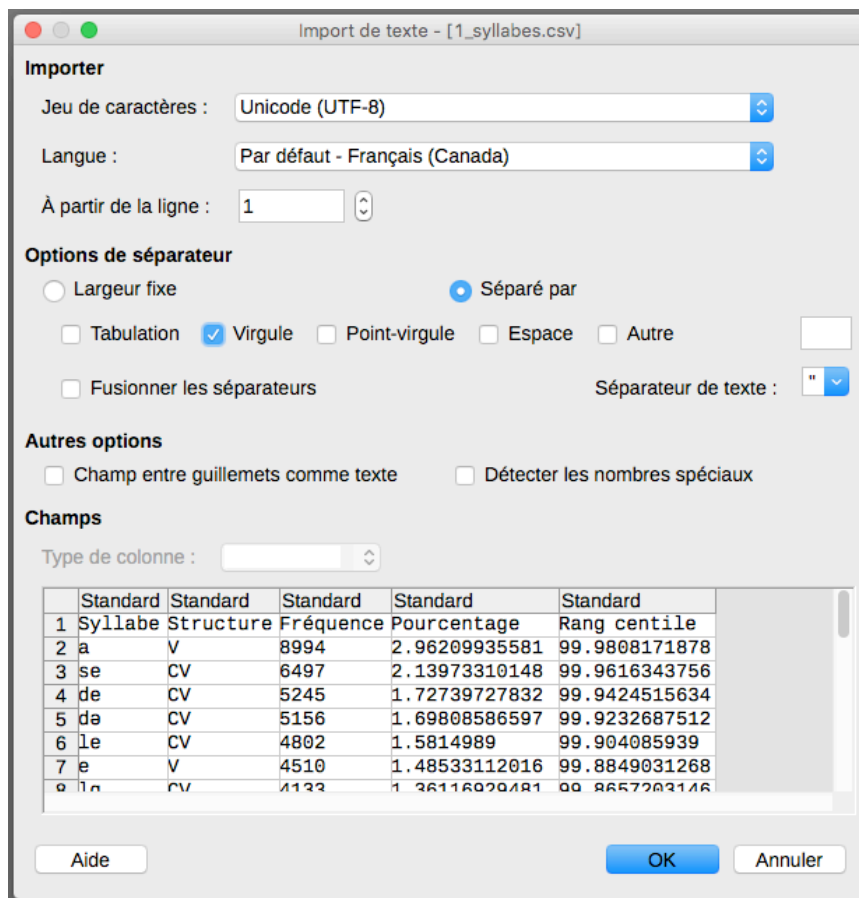
Si LibreOffice est déjà installé dans une autre langue, vous pouvez lui ajouter la votre. Le bouton ci-dessous télécharge uniquement l'interface utilisateur dans votre langue. Si vous ne l'avez pas déjà fait, vous devez d'abord télécharger puis installer le package logiciel principal LibreOffice (ci-dessus).

INTERFACE UTILISATEUR TRADUITE

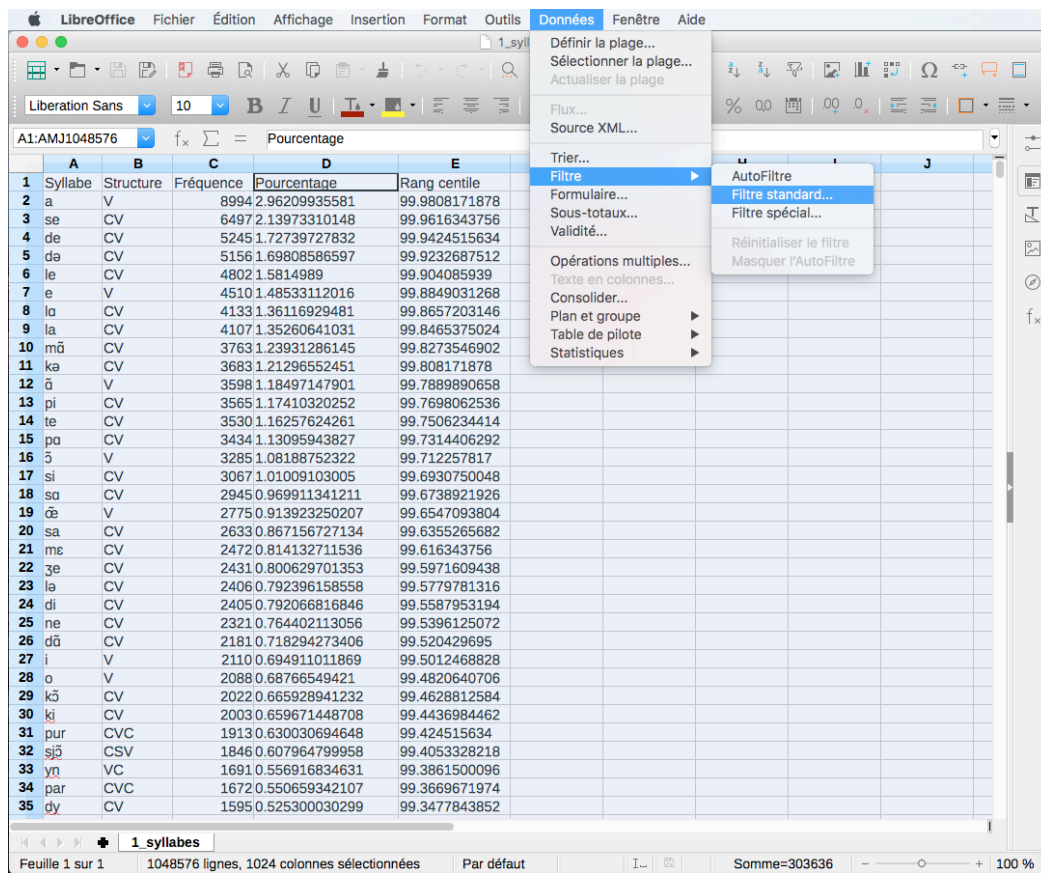
8.8 MB (Torrent, Informations)



7. Sélectionnez **Unicode (UTF-8)** comme jeu de caractères. Sélectionnez **Virgule** comme option de séparateur. Cliquez sur **OK** pour ouvrir.

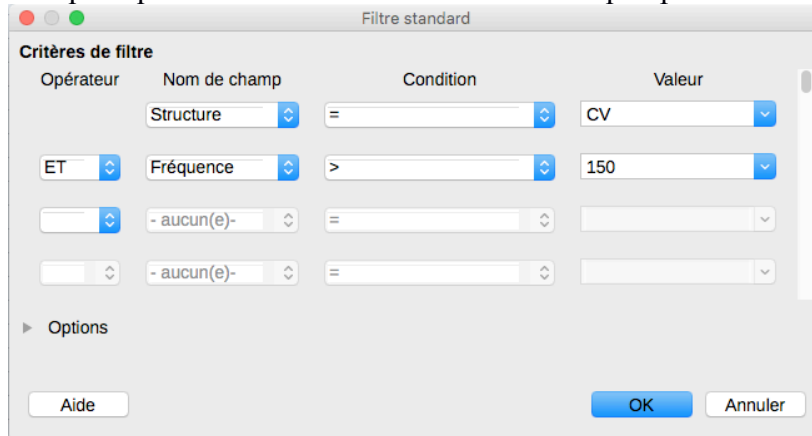


8. Une fois le fichier ouvert, vous pouvez sélectionner les colonnes pertinentes pour votre projet. Masquez ou supprimez les colonnes non désirées.
9. Filtrez les données selon vos besoins : sélectionnez toutes les données, allez au menu, puis cliquez sur **Données – Filtre – Filtre Standard**.



10. Vous pouvez alors utiliser le(s) **filtre(s)** de votre choix pour obtenir l'information désirée.

Indiquez quelle colonne doit être filtrée selon quel paramètre. Par exemple :



11. Vous obtenez vos **résultats!**

Exemple d'élaboration de stimuli

Voici un exemple qui illustre comment des stimuli spécifiques peuvent être construits à partir de *SyllabO+*. À partir du tableau des groupes de 3 syllabes, ces colonnes ont été sélectionnées :

- Triade (transcription en API)
- Structure 1^{ère} syllabe
- Structure 2^e syllabe
- Structure 3^e syllabe
- Rang centile 1^{ère} syllabe
- Rang centile 2^e syllabe
- Rang centile 3^e syllabe

Le filtre « = CV » a été appliqué à toutes les colonnes « structure » et le filtre « >= 95 » a été appliqué à toutes les colonnes « rang centile ». Le tableau résultant de ces filtres (*Tableau 1*) est une sélection de triades de syllabes, où chaque syllabe possède une structure simple « consonne – voyelle » et possède une fréquence élevée puisqu'elle est située entre le 95^e et le 100^e rang centile. Les groupes de 3 syllabes issus de ce tableau peuvent être utilisés comme stimuli dans une étude expérimentale.

Tableau 1. Stimuli: triades de syllabes à structures simples (CV) et fréquence élevée

Triade	Structure 1	Rang centile 1	Structure 2	Rang centile 2	Structure 3	Rang centile 3
kɔ mā se	CV	99.0216765778	CV	99.8273546902	CV	99.9616343756
vu za ve	CV	98.849031268	CV	98.6188375216	CV	98.273546902
mã dɔ ne	CV	99.8273546902	CV	97.602148475	CV	99.5396125072
le zã fã	CV	99.904085939	CV	98.4270093996	CV	97.0650297334
dɪ fe rã	CV	99.5587953194	CV	96.451179743	CV	98.0625359678
...

Voilà ! Bonnes recherches, amusez-vous bien, et restez branchés pour les développements futurs. Merci de citer notre article dans vos publications et rapports découlant de l'utilisation de *SyllabO+* (Bédard et al., 2016) :

Bédard, P., Audet, A.-M., Drouin, P., Roy, J. P., Rivard, J., & Tremblay, P. (2016). SyllabO+: A new tool to study sublexical phenomena in spoken Quebec French. *Behavioral Research Methods*.